

VOUTE-Virtual Overlays Using Tree Embeddings

Stefanie Roos, Martin Beck, Thorsten Strufe
TU Dresden

{stefanie.roos,martin.beck1,thorsten.strufe}@tu.dresden.de

Abstract

Friend-to-friend (F2F) overlays, which restrict direct communication to mutually trusted parties, are a promising substrate for privacy-preserving communication due to their inherent membership-concealment and Sybil-resistance. Yet, existing F2F overlays suffer from a low performance, are vulnerable to denial-of-service attacks, or fail to provide anonymity. In particular, greedy embeddings allow highly efficient communication in arbitrary connectivity-restricted overlays but require communicating parties to reveal their identity. In this paper, we present a privacy-preserving routing scheme for greedy embeddings based on anonymous return addresses rather than identifying node coordinates. We prove that the presented algorithm are highly scalable, with regard to the complexity of both the routing and the stabilization protocols. Furthermore, we show that the return addresses provide plausible deniability for both sender and receiver. We further enhance the routing's resilience by using multiple embeddings and propose a method for efficient content addressing. Our simulation study on real-world data indicates that our approach is highly efficient and effectively mitigates failures as well as powerful denial-of-service attacks.

1 Introduction

Anonymous and censorship-resistant communication is essential for providing freedom of speech. In the last years, threats to this essential human right have emerged in western countries as well, mainly in the form of self-censorship caused by the fear of seemingly private communication being recorded¹. Due to the natural vulnerability of publicly known servers to sabotage and corruption, completely distributed solutions for anonymous communication and content distribution are required. However, the openness of distributed systems presents a vulnerability, enabling attackers to infiltrate the system with a large number of forged participants, as can be seen e.g., from attacks on the Tor [1] network in

¹<http://www.theguardian.com/commentisfree/2013/jun/17/chilling-effect-nsa-surveillance-internet>

2014 ².

F2F overlays circumvent the problem of connecting to permanently changing strangers by restricting connectivity to participants sharing a mutual real-world trust relationship. Hence, adversaries need to resort to social engineering attacks for infiltration of the network. However, large-scale privacy-preserving communication in F2F overlays requires additional measures to achieve anonymity, failure and attack resilience, and efficiency. Multiple studies have shown that deployed F2F overlays such as Freenet [2] are highly inefficient and vulnerable to attacks [3, 4]. Virtual overlays have been proposed as an efficient alternative [3, 5], but recent work has shown that they inherently require unacceptable high stabilization costs [6].

A potential solution is presented by greedy network embeddings such as [7, 8]. Greedy embeddings allow for highly efficient greedy routing in arbitrary connectivity-restricted overlays. For this purpose, they first construct a spanning tree of the network and then assign coordinates based on a node's position in the spanning tree. However, a participant can only contact a non-trusted contact when knowing its coordinate in the network. Though only direct neighbors can directly map the embedding coordinate to a real-world identity, arbitrary participants can reconstruct the social graph based on the revealed coordinates. Participants can then easily be identified from the social graph structure [9] and correlated with their activities in the network due to the coordinate acting as a persistent pseudonym. In this manner, governmental and commercial institutions as well as curious strangers can track individual or all users and establish detailed profiles of their behavior. Possibly even their opinions and interests, published unencrypted in a supposedly anonymous manner, are revealed to the adversary. Thus, network embeddings in their unaltered form fail to provide receiver anonymity.

For utilizing the high efficiency of network embedding, our first requirement is a modified addressing and routing protocol that provides both efficiency and (receiver) anonymity. Second, due to the fragility of spanning trees in the presence of network dynamics and attacks, the resilience of the embedding to both failures and denial-of-service attacks needs to be drastically increased. Rather than only dropping messages, we assume that the adversary first strategically sabotages the embedding algorithm to maximize the impact of its censorship. Third, efficient content storage and retrieval requires the existence of a suitable content addressing scheme for network embeddings.

Our solution addresses the above problems by i) introducing anonymous return addresses to provide receiver anonymity, ii) constructing multiple embeddings and using backtracking during routing to increase the resilience, and iii) utilizing the network embedding to provide a routing protocol for a virtual overlay, thus avoiding the enormous stabilization costs of previous virtual overlays.

Our embedding algorithm assigns coordinates in the form of vectors of b -bit strings, so that nodes in the same subtree of the spanning tree share the same

²<https://blog.torproject.org/blog/tor-security-advisory-relay-early-traffic-confirmation-attack>

prefix, similar to the PIE embedding [8]. Rather than revealing the coordinate, the receiver then generates an anonymous return address by applying a hash cascade to the elements of the coordinate vector salted with a random seed. After publishing the return address and the seed, the receiver can be contacted efficiently without revealing any information that is not required for routing. Furthermore, a node can publish several anonymous return addresses by varying the seed. In this manner, it can construct distinct pseudonyms for distinct contexts, e.g., one pseudonym for each forum discussion it participates in. The revealed information can be further reduced by applying an additional layer of encryption at the price of a reduced efficiency.

By routing in multiple embeddings, we aim to increase the probability of finding a route despite the disruption of routes in some but not all embeddings. We propose a purely local algorithm for the construction of multiple spanning trees of highly different structures to provide largely node-independent routes. In addition, backtracking and a modified distance are integrated into the routing to further improve resilience and avoid congestion.

We evaluate our solution both by a formal security analysis and an extensive simulation study. In the security analysis, we prove a receiver can never be uniquely identified from a return address. Our simulation study indicates that our scheme is highly efficient compared existing approaches in terms of the number of messages required for routing. Furthermore, the resilience is greatly improved. In fact, the routing terminates successfully in the vast majority of cases despite the presence of node failures or powerful attackers, which manipulate the embedding and routing as well as forge connections to honest nodes.

2 Related Work

Here, we describe the state-of-the-art with regard to routing and content discovery in F2F overlays.

The common characteristics of all F2F overlays are i) the restriction of connections to trusted parties, ii) *hop-by-hop anonymization*, i.e., the transfer of messages via a path of trusted nodes that rewrite the source tag of the message to point to themselves and apply probabilistic delays before forwarding a message, iii) encryption of all communication. In the following, we present existing approaches, categorizing them according to their routing methodology in unstructured overlays, virtual overlays, and network embeddings. Routing is applied to either discover nodes based on network coordinates or, more commonly, content based on a content key or description.

Unstructured approaches utilize *flooding*, e.g. in Turtle [10], or *probabilistic forwarding* e.g. in OneSwarm [11]. GnuNet attempts to combine random walks with deterministic routing [12]. These overlays focus locating content rather than individual nodes. Due to the replication of content, the content can indeed be located, but efficient communication between two uniquely defined entities is not possible.

Virtual overlays address the problem of establishing an overlay despite the

restricted connectivity by replacing overlay links with tunnels of trusted nodes. So, efficient tunnel discovery and maintenance is a main concern given the inherent network dynamics: Vasserman et al. [3] suggest flooding the network for discovering adequate overlay neighbors, thus creating a large overhead. In contrast, X-Vine leverages the overlay routing by concatenating previously existing tunnels to a new one, thus entailing a increase of the average tunnel length and hence routing costs over time [5]. Indeed, without an additional routing protocol in the underlying F2F overlay, efficient maintenance and efficient routing are inherently mutually exclusive in a virtual overlay [6].

In contrast, *network embeddings* assign coordinates that allow efficient routing to nodes. For example, the F2F mode of Freenet relies on a network embedding. However, results indicate that the embedding is lacking both with regard to routing efficiency [3] and attack resilience [4]. Hoefer et al. [13] propose highly efficient *greedy embeddings*. However, their approach reveals the identity of the communicating parties and fails to consider resilience. Furthermore, their proposed scheme for content addressing maps the majority of content keys to the same central node.

In summary, network embeddings are the only existing approach providing a high efficiency at acceptable maintenance costs. However, achieving receiver anonymity, resilience, and suitably content addressing is an unsolved highly challenging problem.

3 Adversary Model

We aim to realize efficient F2F overlays making use of network embeddings but at the same time providing receiver anonymity, resilience, and content addressing. Note that we do not consider sender anonymity because the problem of sender anonymity can easily be solved by starting the routing with a short random walk, as extensively analyzed for various anonymous look-up strategies for distributed hash tables (DHTs) (e.g., [12, 14]). In contrast, receiver anonymity is a challenging problem for network embeddings, because the coordinates acting as a node's pseudonym are essential for the routing process and hence for the efficient communication between arbitrary node pairs. The term resilience is loosely defined. In general, a system is denoted resilience if an action is only slightly impaired by node failures or attacks. Commonly, a system is judged to be resilient by comparison with others.

We consider two attack goals in our adversary model. The first goal is to discover the identity of communicating parties, in particular the identity of the designated receiver of a message. A second goal of the attacker is to block undesired communication using a so-called *black hole attack* [15], which could be applied in case an attacker fails to identify specific parties. During such an attack, an adversary indiscriminately censors communication by first gaining a predominant position in the system and then dropping all received messages. In addition, attacks on the availability, such as pollution, i.e., denial-of-service attacks by flooding the network with content and traffic, and eclipse attacks, i.e.,

censoring of specific content, present a threat for any P2P systems. However, these attacks have been addressed in various publications (e.g., [15]), which can be applied to our contribution with few modifications. Hence, we do not consider them in our evaluation.

As for the attacker’s capacities, we assume a local, active, internal, possibly colluding attacker, able to drop and manipulate messages it receives. The adversary can control one or several colluding nodes in the network but is unable to observe the complete topology. In particular, we assume that an adversary cannot be certain that it knows all neighbors of a node, in other words, the complete circle of a user. We assume that this is hard, because it requires the adversary to i) be sure that he knows all contacts from different social circles of a user, such as family, close friends, and colleagues, and ii) establish connections to all of them. A global passive attacker is disregarded on the basis that steganographic techniques can be applied to hide the F2F traffic as suggested in e.g. [16]. However, attackers are modeled as polynomial time adversaries, which are given a transcript of all own and public input, as well as all locally observed traffic. Their computation power is therefore bounded by polynomial time algorithms, which prevents breaking computationally-secure cryptographic primitives. We assume that an adversary can easily forge an arbitrary number of nodes, so called *Sybils*. However, gaining connections and hence influence in a F2F overlay requires establishing real-world trust relationships. Such *social engineering attacks* are considered to be costly and difficult because they require long-term interaction between a human adversary and an honest participant. Thus, the number of connections between honest nodes and forged participants can be assumed to be small. More precisely, we assume that the number is logarithm with the network size, in agreement to previous work [5].

4 Network Embeddings

Our solution builds upon previous work in the area of network embeddings, which assign coordinates to nodes with the goal of structuring networks, e.g., for efficient routing in wireless sensor networks or as an alternative to the current IP layer in content-centric networking. We first introduce some notation, then explain the principal concepts of network embeddings, and conclude by describing specific algorithms. In particular, we detail the PIE embedding [8], which we modify in Section 5 to allow for anonymity.

4.1 Basic Terminology

In the remainder of the paper, we represent an overlay network by a graph $G = (V, E)$ with *nodes* V and *links* or *edges* $E \subset V \times V$. Because we require mutual trust for connection establishment in F2F overlays, the network is bidirectional. We denote the neighbors of u by $N_u^G = \{v \in V : (u, v) \in E\}$. Therefore, a *Friend-to-friend (F2F) overlay* is an overlay such that the set of links is given by pairs of nodes sharing a real-world trust relationship. Embedding algorithms

heavily rely on *spanning trees*, connected subgraphs $ST = (V, E^T)$ of G such that $|E^T| = |V| - 1$. In such a (spanning) tree, one node r is designated as the *root* and the position of nodes are described based on their relation to the root. In particular, the *level* or *depth* of a node u is given by the length of the path from u to r . If u is not the root, the *parent* of u is defined to be a neighbor $v \in N_u^{ST}$ with a shorter path to r than u , whereas the remaining neighbors are u 's *children*. A node without children is called a *leaf*, whereas nodes with children are called *internal nodes*.

4.2 Concept

Now, we define the concept of network embeddings and in particular *greedy* network embeddings. In the following, let $G = (V, E)$ be a network and (\mathbf{X}, δ_X) be a metric space with a distance δ_X . A network embedding is defined as a function $id : V \rightarrow \mathbf{X}$ assigning each node a coordinate. The problem of enabling routing in a connectivity-restricted network has been addressed by the design of *greedy embeddings*. Greedy embeddings [17] are coordinate assignments, such that for any source-destination pair $(s, t) \in V \times V$ with $s \neq t$, a neighbor u of s exists such that $\delta_X(u, t) < \delta_X(s, t)$. We say that u is closer to t than s with regard to δ_X . As a consequence, straight-forward greedy routing is guaranteed to find a route from s to t .

Though there exists a multitude of greedy embedding algorithms, they all follow the same four abstract steps: i) Construct a spanning tree T , ii) Each internal node in T enumerates its children, iii) The root receives a predefined coordinate, iv) Children derive their coordinate from the parent's coordinate and the enumeration index assigned by the parent (e.g. [7, 8, 18, 19]). The coordinates are then distributed such that the embedding of the spanning tree is greedy, as specified for the PIE embedding below. Subsequent to the coordinate assignment, nodes consider all neighbors, including those that are neither parent nor child, for the routing. So, routing is not restricted to tree edges. We call non-tree edges *shortcuts* because they allow for a faster reduction of the distance and shorter routes than predicted by the distance in the tree.

In the following, we consider the construction and stabilization costs for such greedy embeddings. A spanning tree is constructed by i) selecting a root node using a distributed leader election protocol such as [20, 21], and ii) building the tree from the root. In this manner, it is possible to construct a spanning tree with $\mathcal{O}(n \log n)$ messages for a graph of diameter $\mathcal{O}(\log n)$ [20], though integrating protections against nodes aiming to cheat the root selection protocol such as [21] require a higher cost. Various embeddings [8, 18, 19] are able to react to dynamics without computing the complete embedding whenever the topology changes. New nodes join the trees as leaves, requiring only a constant overhead for contacting one of their neighbors to be their parent and receiving a coordinate from said parent. If any node but the root leaves, only its descendants have to reconnect. We show that the stabilization overhead then scales linearly with the tree depth rather than linear with the number of participants.

4.3 Existing Approaches

Though embedding algorithms generally rely on a spanning tree and assign coordinates according to the tree structure, the nature of the assigned coordinates is highly diverse: Embeddings into hyperbolic space such as [7, 18, 19] allow embedding in low-dimensional spaces. However, proposed hyperbolic embeddings are extremely complex and do not scale with regard to the number of bits required for coordinate representation [19]. Custom-metric approaches have been designed to overcome these shortcomings. The custom-metric embedding PIE [8] assigns an empty vector as the root coordinate. Child coordinates are then derived from the parent coordinate by concatenating the parent coordinate with the index assigned to the child by the parent, potentially weighted with the cost of the parent-child edge if such weights are given. In this manner, a node s 's coordinate represents the route from the root to u . Consequently, the distance δ_X is given by the hop distance of two nodes in the tree. An example for the PIE embedding in unweighted graphs is displayed on the left side of Figure 1. Whereas routing in greedy embeddings is highly efficient in comparison to non-greedy embeddings [13], neither anonymity nor resilience has been considered in suitably manner.

5 Design

Our main contribution lies in proposing multiple greedy embeddings with anonymous return addresses and a virtual overlay on top of the embeddings. In the following, we present our system, in particular

- a spanning tree construction and stabilization algorithm for multiple parallel embeddings,
- an embedding algorithm providing efficiency as well as allowing for improved censorship-resistance through a modified distance,
- an address generation algorithm **AdGen**_{node} enabling receiver anonymity, and
- a virtual overlay design based on embeddings which allowing balanced content distribution and efficient content retrieval.

5.1 Tree Construction and Stabilization

In this section, we show how we construct and stabilize γ parallel spanning trees. In the next section, we then describe how to assign coordinates on the basis of these trees. We want to increase the robustness and censorship-resistance by using multiple trees. In order to ensure that the trees indeed offer different routes, our algorithm encourages nodes to select different parents in each tree if possible. Our algorithm design follows similar principles as the provable optimally robust and resilient tree construction algorithm for P2P-based video streaming

presented in [22]. However, the algorithm assumes that nodes can change their neighbors. Thus, we cannot directly apply the algorithm nor the results. In the following, we first discuss the tree construction and then the stabilization.

Tree Construction: We divide the construction of a tree into two phases i) selecting the root, and ii) building the tree starting from the root. We can apply [20] for the root election, which achieves a communication complexity of $\mathcal{O}(n \log n)$. Our own contribution lies in ii) the tree construction after the root node has been chosen.

We now shortly describe the idea of our algorithm and then the actual algorithm. A node u that is not the root receives messages from its neighbors when they join a tree and are hence available as parent nodes. There are two questions to consider when designing an algorithm governing u 's reaction to such messages, called invitations in the following. First, u has to decide if and when it accepts an invitation. Second, u has to select an invitation in the presence of multiple invitations.

For the second question, u always prefers invitation from nodes that have been their parent in less trees with the goal of constructing different trees and increasing the overall number of possible routes. Increasing the number of routes allows the use of alternative routes if the request can not be routed along the preferred route due to a failed or malicious node. If two neighbors are parents in the same number of trees, u can either select one randomly or prefer the parent closer to the root. Choosing a random parent reduces the impact of nodes close to the root but is likely to lead to longer routes and thus a lower efficiency.

Coming back to the first question of if and when u accepts invitations, u should always accept an invitation of a neighbor v that is not yet a parent of u in any tree in order to choose different parents as often as possible. In contrast, if v is already a parent, u might wait for the invitation of a different neighbor. However, it is unclear if it is possible for all neighbors of u to ever become a parent. For example, a neighbor of degree 1 is only a parent if it is the root. In order to overcome this dilemma, u periodically probabilistically decides if it should accept v 's invitation or wait for another invitation. So, u eventually accepts an invitation but does provide alternative parents the chance to send an invitation.

Now, we describe the exact steps of the algorithm. The algorithm is a round-based invitation protocol for the tree construction. After a node u is included in the i -th tree, u sends invitations (i, u) to all its neighbors inviting them to be its children in tree i . When u receives an information (j, w) for the j -th tree from a neighbor w , it saves the invitation if it is not yet contained in tree j and otherwise stores it. The invitation can still be used if u has to modify its parent selection later. In each round, a node u considers all invitations for trees it is not yet part of, as described in Algorithm 1. Let $pc(v)$ be number of trees for which a neighboring node v is a parent of u . If u has received invitations from neighbors v with minimal $pc(v)$ among all neighbors, u accepts one of those invitations (Lines 1-3). In the presence of multiple invitations,

we experiment with two selection strategies: i) Choosing a random invitation, and ii) Choosing a random invitation from a node on the lowest level. The latter selection scheme requires that the invitations also detail the level of the potential parent node in the tree. If u does not have an invitation from any node with minimal $pc(v)$, u nevertheless accepts an invitation with probability q in order to guarantee the termination of the tree construction. If u accepts a parent, it selects a node v that has offered an invitation and has the lowest $pc(v)$ among neighbors with outstanding invitations (Lines 7-8). In this manner, we guarantee the convergence of tree construction.

The acceptance probability q is essential for the diversity and the structure of the trees: For a high q , nodes quickly accept invitations leading to trees of a low depth and thus short routes. However, in the presence of an attacker acting as the root of all or most trees, the trees are probably close to identical, resulting in a low censorship-resistance. A lower acceptance probability q increases the diversity but entails longer routes. Thus, a low q results in a higher communication complexity and at some point decreases the robustness due to the increased likelihood of encountering failed nodes on a longer route. In Section 6.1, we show that the constructed trees are of a logarithmic depth such that we indeed maintain a routing complexity of $\mathcal{O}(\log n)$. Note that Algorithm 1 does not assume that all trees are constructed at the same time. Rather, individual trees can be (re-)constructed while the remaining trees impact the parent choice in the new tree but remain unchanged.

Algorithm 1 `constructTreeRound()`

{Internal state: Set I of invitations, acceptance probability q , $pc : N_u \rightarrow \mathbb{N}_0$ number of times neighbor is parent}

```

1:  $PP \leftarrow \{(i, w) \in I : \forall \mathbf{v} \in \mathbf{N}_u : pc(w) \leq pc(v)\}$ 
2: if  $PP$  is not empty then
3:   Select invitation in  $PP$  to answer
4: else
5:    $r \leftarrow$  uniform random number
6:   if  $r \leq q$  then
7:      $PQ \leftarrow \{inv = (i, w) \in I : \forall (\mathbf{j}, \mathbf{v}) \in \mathbf{I} : pc(w) \leq pc(v)\}$ 
8:     Select invitation in  $PQ$  to answer
9:   end if
10: end if
```

Stabilization: Now, we consider the stabilization of the trees when nodes join and leave. Stabilizing the trees efficiently, i.e., repairing them locally rather than reconstructing the complete tree whenever the topology changes, is essential for efficiency. Joining nodes can be integrated in a straight-forward manner by connecting to their neighbors as children, again trying to maximize the diversity of the parents. For this purpose, nodes record the time, i.e., the round in our abstract time model, they joined the tree. Now, when a new node u joins, it

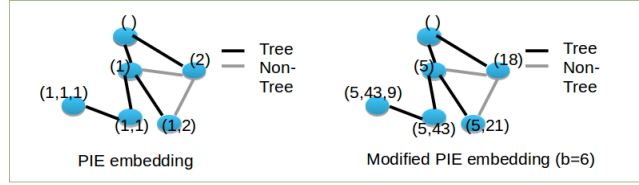


Figure 1: Original PIE and modified PIE coordinates using $b = 6$ -bit number

requests its neighbors' coordinates and these timestamps for all trees. Based on this information, u can simulate Algorithm 1 locally, ensuring that its expected depth in the tree is unaffected by its delayed join. When a node departs, all its children have to choose a different parent and inform their descendants of the change. In order to prevent a complete subtree from being relocated at an increased depth, the descendants may also select a different parent. The selection of the new parent again follows Algorithm 1 but only locally re-establishes the trees affected by the node departure. We show that the stabilization complexity considering any node but the root is linear in the terms of average depth of the node in the trees in Section 6.1.

We formally prove that the above stabilization algorithm indeed only introduces only logarithmic complexity in Section 6.1. We now present the embedding algorithm, which in agreement with the presented tree construction algorithm, assigns coordinates within subtrees independently of the remaining subtrees to allow for local stabilization.

5.2 Embedding and Routing

In this section, we show how we assign coordinates in a spanning tree and how to route based on these coordinates. As we want to prevent an attacker from guessing the coordinate of a receiver, we require a certain degree of in-determinism in the coordinate assignment. We thus choose a slightly modified version of the unweighted PIE embedding [8], which we have introduced in Section 4. Our main modification lies the use of in-deterministic coordinates in order to prevent an adversary from guessing the coordinate and thus undermining the anonymization schemes presented in the next section. The routing algorithm corresponds to the greedy routing with backtracking. In addition to the tree distance in [8], we also present a second distance preferring nodes with a long common prefix and thus avoiding routes via nodes close to the root whenever possible. In this manner, we increase robustness and censorship-resistance because the routing algorithm considers alternative routes and the impact of strategically choosing a position close to the root is reduced. In the following, we subsequently present the embedding algorithm, the distance functions, and the routing with backtracking.

Embedding Algorithm: Embeddings are performed on each of the γ trees independently, so that we only consider one embedding id . Throughout this section, let b be a sufficiently large integer, $PRNG$ a pseudo-random number generator with values in \mathbb{Z}_2^b , and $h : \{0, 1\}^* \rightarrow H$ a cryptographically secure hash function. We describe the embedding algorithm, then the distance used for routing, and last, the backtracking procedure, which allows highly resilient routing despite failures.

We now describe the embedding algorithm for one tree. The coordinate assignment starts at the root and then spreads successively throughout the tree. After a spanning tree has been established, the root r is assigned an empty vector as a coordinate $id(r) = ()$. In the next step, each child v of the root generates a random b -bit number $a \in \mathbb{Z}_2^b$ such that its coordinate is $id(v) = (a)$. Here, our algorithm differs from the PIE embedding because it uses random rather than consecutive numbers, thus preventing an adversary from guessing the coordinate in an efficient manner. Subsequently, nodes in the tree are assigned coordinates by concatenating their parent's coordinate with a random number. So, upon receiving its parent coordinate $id(p(v)) = (a_1, \dots, a_{l-1})$, a node v on level l of the tree obtains its coordinate $id(v) = (a_1, \dots, a_{l-1}, a_l)$ by adding a random b -bit number a_l . The coordinate space is hence given by all vectors consisting of b -bit numbers, i.e., $\mathbf{X} = \{(a_1, \dots, a_{l-1}, a_l) : l \in \mathbb{N}_0, a_i \in \{0, 1\}^b\}$. Figure 1 displays the differences between the original PIE embedding and our variation.

Note that the independent random choice of the b -bit number $a \in \mathbb{Z}_2^b$ might lead to two nodes having the same coordinate. Thus, b should be chosen such that the chance of equal coordinates should be negligible. If two children nevertheless select the same coordinate, the parent node should inform one of them to adapt its choice. Note that allowing the parent to influence the coordinate selection in this manner does not really increase the vulnerability to attacks, as the parent can achieve at least the same damage by constantly changing its coordinate. Such constant changes can be detected easily, so that nodes should stop selecting such nodes as parents. In general, by moving the choice of the last coordinate element from the parent to the child, we automatically reduce the impact of a malicious parent as it can not determine the complete coordinate of the child.

Distances: We still need to define distances between coordinates in order to apply greedy routing. For this purpose, we consider two distances on \mathbf{X} . Both rely on the common prefix length $cpl(x_1, x_2)$ of two vectors x_1 and x_2 and the coordinate length $|x_1|$.

First, we consider the tree distance δ_{TD} from [8], which gives the length of path between the two nodes in the tree, i.e.,

$$\delta_{TD}(x_1, x_2) = |x_1| + |x_2| - 2cpl(x_1, x_2). \quad (1)$$

Secondly, the common prefix length can be used as the determining factor in the distance function, i.e., for a constant L exceeding the length of all node

request. For this purpose, all nodes remember their predecessor on the routing path as well as the neighbors they have forwarded the request to. If all neighbors closer to the target have been considered and have been unable to deliver the request, the node reroutes the request to its predecessor for finding an alternative path. The routing is thus only considered to be failed if the request returns to its source s and cannot be forwarded to any other neighbor. In this manner, all *greedy paths*, i.e., all paths with a monotonously decreasing distance to the target, are found.

Algorithm 2 route()

```

    {Input: current node  $u$ , message  $msg$  from node  $w$ , tree index  $i$ , target
    coordinate  $x_i$ }
    {Internal state: set  $S(msg)$  of nodes  $u$  forwarded mess to, predecessor
     $pred(msg)$ , distance  $\delta$ }
1: if  $id_i(u) == x_i$  then
2:   Routing succeeds
3: else
4:   {Store predecessor unless backtracking}
5:   if not  $S(msg)$  contains  $w$  then
6:      $pred(msg) \leftarrow w$ 
7:   end if
8:   {Determine closest neighbors}
9:    $C \leftarrow \operatorname{argmin}_{v \in N_u \setminus S(msg)} \delta(id_i(v), x_i)$ 
10:   $next \leftarrow$  random element in  $C$ 
11:  if  $\delta(id_i(v), x_i) > \delta(id_i(next), x_i)$  then
12:    Forward  $msg$  to  $next$  {Forward if improvement}
13:  else
14:    if  $pred(msg)$  is set then
15:      Forward  $msg$  to  $pred(msg)$  {Backtrack}
16:    else
17:      Routing failed
18:    end if
19:  end if
20: end if

```

Algorithm 2 gives the pseudo code describing one step of the routing algorithm, including the backtracking procedure. When receiving a message msg , the node u first checks if it is the receiver of msg , thus successfully terminating the routing (Line 2). If u is not the receiver, it determines if the routing is currently in the backtracking phase by checking if u has previously forwarded msg to the sender w . Otherwise, it stores the sender of msg as a predecessor for potential later backtracking (Line 5). In the manner of greedy routing, u selects the closest neighbor to the target coordinate. In the presence of several closest neighbors, u picks one of them uniformly at random (Lines 7-8). Note that in the presence of failures, the embedding can lose its greediness. Hence, to avoid loops, u only forwards the request to that neighbor if it is indeed closer (Line 10). Otherwise, u contacts its predecessor (Line 13) or forfeits the routing

if no such predecessor exists (Line 15), i.e., if u is the source of the request.

This completes the description of the routing and stabilization functionalities. However, up to now, we used identifying coordinates rather than anonymous addresses.

5.3 Anonymous Return Addresses

In this section, we introduce our address generation algorithm for generating anonymous return addresses but do not reveal the receiver of the request. For this reason, we call the generated addresses *route preserving (RP) return addresses*. Based on these return addresses, we specify two routing algorithms \mathbf{R}^{TD} and \mathbf{R}^{CPL} for routing a request containing a return address. The return addresses allow a node to determine the common prefix length of their neighbor's coordinates and the receiver coordinate, which allows the node to determine the closest neighbor. Hence, \mathbf{R}^{TD} and \mathbf{R}^{CPL} correspond to Algorithm 2 for the two distance function δ_{TD} and δ_{CPL} when using return addresses rather than a receiver coordinates. After describing the algorithm, we show that the return addresses indeed preserve routes.

Return Address Generation: Return addresses are generated in three steps:

1. Padding the coordinate
2. Applying a hash cascade to obtain the return address
3. Adding a MAC

Algorithm 3 displays the pseudo code of the above steps.

Algorithm 3 generateRP()

{Input: coordinate $x = (a_1, \dots, a_l)$, seed s, s_{pad} }
 {Internal State: key $\mathbb{K}_{MAC}(v)$, h , PRNG}

```

1:  $\tilde{k} \leftarrow PRNG(s)$ 
2:  $d_1 \leftarrow h(\tilde{k} \oplus a_1)$ 
3: for  $j = 2 \dots L$  do
4:   if  $j \leq l$  then
5:      $a'_j \leftarrow a_j$ 
6:   else
7:      $a'_j \leftarrow PRNG(s_{pad} + j)$  {Padding}
8:   end if
9:    $d_j \leftarrow h(d_{j-1} \oplus a'_j)$  {Hash cascade}
10: end for
11:  $mac \leftarrow h(\mathbb{K}_{MAC}(v) || d_1 || d_2 || \dots || d_L)$  {MAC}
12: Publish  $y = (d_1, \dots, d_L), \tilde{k}, mac$ 
```

The first step of the return address generation prevents an adversary from identifying coordinates based on their length. A node v pads its coordinate

$x = (a_1, \dots, a_l)$ by adding random elements a'_{l+1}, \dots, a'_L . More precisely, v selects a seed s_{pad} for the pseudo-random number generator $PRNG$ and obtains the padded coordinate $x' = (a'_1, \dots, a'_l, a'_{l+1}, \dots, a'_L)$ with

$$a'_j = \begin{cases} a_j, & j \leq l \\ PRNG(s_{pad} \oplus j), & j > l \end{cases}.$$

In order to ensure that the closest node coordinate to x' is indeed x , v recomputes the padding with a different seed if a'_{l+1} is equal to the $l + 1$ -th element of a child's coordinate³. Afterwards, v chooses a different seed s for the construction of the actual return address and generates $\tilde{k} = PRNG(s) \in \tilde{\mathbb{K}} = \mathbb{Z}_2^b$. v then executes the local function $hc : \mathbf{X} \rightarrow \mathbf{Y} = H^L$ in order to obtain a vector y with elements in H . The i -th element of $y = (d_1, \dots, d_L)$ is given by

$$d_j = \begin{cases} h(\tilde{k} \oplus a'_1), & j = 1 \\ h(d_{j-1} \oplus a'_j), & j = 2 \dots L \end{cases}. \quad (3)$$

We call the pair (y, \tilde{k}) a *return address*, which can be used to find a route to the node with coordinate x . Before publishing the return address, v adds a MAC $mac(y_i, \mathbb{K}_{MAC}(v)) = h(d_1 || \dots || d_L || \mathbb{K}_{MAC}(v))$ for a private key $\mathbb{K}_{MAC}(v)$ to prevent malicious nodes from faking return addresses and gaining information from potential replies. Last, v publishes the return address (y, \tilde{k}) and the MAC.

Routing Algorithms: Now, we determine diversity measures $\delta_{RP-TD} : \mathbf{X} \times \mathbf{Y} \rightarrow \mathbb{R}_+$ and $\delta_{RP-CPL} : \mathbf{X} \times \mathbf{Y} \rightarrow \mathbb{R}_+$ in order to compare coordinates x and y with regard to δ_{TD} and δ_{CPL} . The diversity measure then assumes the role of the distance δ in Algorithm 2.⁴

In order to define a sensible diversity measure, note that for any coordinate c and return address y for coordinate x , we have $cpl(x, c) = cpl(y, hc(c, \tilde{k}))$. We thus can define the diversity measure in terms of the common prefix length in the same manner as the distance. More precisely, for $* \in \{TD, CPL\}$, the diversity $\delta_{RP-*}(y, \tilde{k}, c)$ for of a coordinate c to the return address y is

$$\delta_{RP-*}(y, \tilde{k}, c) = \delta_*(y_i, hc(c, \tilde{k})). \quad (4)$$

In practice, u can increase the efficiency of the computation by only determining $hc(c, \tilde{k})$ up to the first element in which it disagrees with y . Thus, we now have two possible realizations of the routing algorithm \mathbf{R}_{node} , namely \mathbf{R}^{TD} and \mathbf{R}^{CPL} . Given the RP return address (y, \tilde{k}) of the destination e , \mathbf{R}^{TD} and \mathbf{R}^{CPL} forward the message to the neighbor v with the lowest diversity measure $\delta_{RP-TD}(y, \tilde{k}, id(v))$ and $\delta_{RP-CPL}(y, \tilde{k}, id(v))$, respectively.

³We exclude this step in Algorithm 3 for increased readability

⁴Note that a diversity measure is not a distance because it i) is defined for two potentially distinct sets \mathbf{X} and \mathbf{Y} , and ii) is not symmetric.

Proving Route Preservation: We now prove formally that the above return addresses preserve routes. For this purpose, we first define the notion of preserving a property of a coordinates. Note that we

Definition 5.1. Let $Q_u : \mathcal{P}(\mathbf{X}) \times \mathbf{X} \rightarrow \mathcal{P}(\mathbf{X})$ be a local function of node u in a graph $G = (V, E)$. Given a set $C \subset \mathbf{X}_V = \{v \in V : id(v)\}$ of node coordinates and a target coordinate $x \in \mathbf{X}$, Q_u returns a subset $C' \subset C$. A return address (y, \tilde{k}) for a coordinate x is said to preserve Q if for all $u \in V$, there exists a function $Q'_u : \mathcal{P}(\mathbf{X}) \times \mathbf{Y} \times \tilde{\mathbb{K}} \rightarrow \mathcal{P}(\mathbf{X})$ such that for all $C \subset \mathbf{X}$

$$Q'_u(C, y, \tilde{k}) = Q_u(C, x).$$

The notion of *route preserving (RP)* return addresses now follows if we choose the function Q_u to return the neighbors with the closest coordinates to $cord(y, \tilde{k})$.

Definition 5.2. Let

$$\begin{aligned} ra_u : \mathcal{P}(\mathbf{X}) \times \mathbf{X} &\rightarrow \mathcal{P}(\mathbf{X}), \\ ra_u(C, x) &= argmin_{c \in C} \{\delta(c, x)\} \end{aligned} \tag{5}$$

determine the closest coordinates in a set C to a coordinate x . A return address (y, \tilde{k}) is called *route preserving (RP)* (with regard to δ) if it preserves ra .

Based Definition 5.2, we can now show that Algorithm 3 generates RP return addresses.

Theorem 5.3. Algorithm 3 generates RP return addresses with regard to the distances δ_{TD} and δ_{CPL} .

Proof. In order to show that (y, \tilde{k}) preserves routes, we derive the relation between the diversity measures δ_{RP-TD} and δ_{RP-CPL} , defined in Eq. 4, and the corresponding distances δ_{TD} and δ_{CPL} , defined in Eq. 1 and Eq. 2, respectively.

Let $cord(y, \tilde{k})$ denote the padded coordinate used to generate y , and let x be the coordinate without padding. In the following, we relate the distance of x and a coordinate c to the diversity measure of (y, \tilde{k}) and c . We can assume that $cpl(cord(y, \tilde{k}), c) = cpl(x, c) \leq |x|$, i.e., the common prefix length of the padded coordinate and c is at most equal to the length of the original coordinate x . A node with coordinate c with $cpl(cord(y, \tilde{k}), c) > |x|$ cannot exist in a valid embedding. More precisely, our embeddings algorithm ensures that coordinates are unique and a node v ensures that the first element of the padding does not corresponds to the $|id(v)| + 1$ -th element of a descendant's coordinate. Thus, the coordinate x is the unique closest coordinate of a node to the padded coordinate. Thus, we can indeed limit our evaluation to coordinates c with $cpl(cord(y, \tilde{k}), c) \leq |x|$.

We start by considering the tree distance δ_{TD} . By Eq. 4, we have

$$\begin{aligned} \delta_{RP-TD}(y, \tilde{k}, c) &= L + |c| - 2cpl(cord(y, \tilde{k}), c) \\ &= |x| + |c| - 2cpl(x, c) + (L - |x|) \\ &= \delta_{TD}(x, c) + (L - |x|). \end{aligned}$$

Hence, diversity measure and distance only differ by a constant independent of c . Thus, any forwarding node can determine the closest coordinates to the destination in its neighborhood and thus Algorithm 3 generates RP return addresses with regard to δ_{TD} .

For the distance δ_{CPL} , we consider two coordinates c_1 and c_2 with $cpl(cord(y, \tilde{k}), c_i) = cpl(x, c_i)$ for $i = 1, 2$. We show that i) $\delta_{CPL}(x, c_1) = \delta_{CPL}(x, c_2)$ iff $\delta_{RP-CPL}(y, \tilde{k}, c_1) = \delta_{RP-CPL}(y, \tilde{k}, c_2)$ and ii) $\delta_{CPL}(x, c_1) < \delta_{CPL}(x, c_2)$ iff $\delta_{RP-CPL}(y, \tilde{k}, c_1) < \delta_{RP-CPL}(y, \tilde{k}, c_2)$. Thus, the return address (y, \tilde{k}) is RP as the comparison of two coordinates yields the same order when using the return address as for the original coordinate. For i) note that by Eq. 2 $\delta_{CPL}(x, c_1) = \delta_{CPL}(x, c_2)$ implies that $cpl(x, c_1) = cpl(x, c_2)$ and $|c_1| = |c_2|$. Because $cpl(cord(y, \tilde{k}), c_i) = cpl(x, c_i)$, we have $\delta_{RP-CPL}(y, \tilde{k}, c_1) = \delta_{RP-CPL}(y, \tilde{k}, c_2)$. The converse holds analogously by Eq. 4. If ii) $\delta_{CPL}(x, c_1) < \delta_{CPL}(x, c_2)$, then Eq. 2 implies that either $cpl(x, c_1) > cpl(x, c_2)$ or $cpl(x, c_1) = cpl(x, c_2)$ and $|c_1| < |c_2|$. In the first case, the claim follows as $cpl(cord(y, \tilde{k}), c_i) = cpl(x, c_i)$ and δ_{CPL} and δ_{RP-CPL} both prefer coordinates with a longer common prefix length. For the second case, the claim follows under the assumptions $cpl(x, c_1) = cpl(x, c_2)$ and $cpl(cord(y, \tilde{k}), c_i) = cpl(x, c_i)$, because

$$\begin{aligned}
& \delta_{CPL}(x, c_1) < \delta_{CPL}(x, c_2) \\
& \iff L - cpl(x, c_1) - \frac{1}{|x| + |c_1| + 1} < L - cpl(x, c_2) - \frac{1}{|x| + |c_2| + 1} \\
& \iff -\frac{1}{|x| + |c_1| + 1} < -\frac{1}{|x| + |c_2| + 1} \\
& \iff |x| + |c_1| + 1 < |x| + |c_2| + 1 \\
& \iff L + |c_1| + 1 < L + |c_2| + 1 \\
& \iff -\frac{1}{L + |c_1| + 1} < -\frac{1}{L + |c_2| + 1} \\
& \iff \delta_{CPL}(cord(y, \tilde{k}), c_1) < \delta_{CPL}(cord(y, \tilde{k}), c_2) \\
& \iff \delta_{RP-CPL}(y, \tilde{k}, c_1) < \delta_{RP-CPL}(y, \tilde{k}, c_2)
\end{aligned}$$

Hence for both cases i) and ii), Algorithm 3 generates RP return addresses with regard to δ_{CPL} . \square

Up to now, we have only considered route preserving return addresses generated by padding coordinates and applying a hash cascade. Optionally, an additionally layer of symmetric encryption can be added, preventing a node v from deriving the actual length of the common prefix. Rather, v can only determine if a neighbor is closer to the destination than v itself. However, we show the same degree of anonymity for for both algorithms, so that the additional layer does not result in a provably higher level of anonymity. Furthermore, the additional layer reduces the efficiency as nodes select one closer neighbor at random rather than the closest neighbor. For this reason, the advantage of the additional layer is limited, so that we focus on RP return addresses here and defer the further obfuscation of coordinates to the appendix.

We prove that Algorithm 3 indeed enables receiver anonymity in Section 8.

5.4 Content Storage

In order to store content, we use a distributed hash table (DHT). As nodes can not communicate directly, they store tree addresses in their routing tables and leverage the tree routing. In this manner, we do not require maintenance-intensive tunnels like [5] and [3]. Note that we only sketch the solution for content storage and retrieval because our focus lies in improving the quality of the greedy embeddings for messaging between nodes. In the following, we first present the idea of our design and then a realization based upon a recursive Kademlia.

General Design: Nodes establish a DHT by maintaining a routing table of (virtual) overlay connections. The routing table contains entries correspond to a DHT coordinate and corresponding return addresses. Nodes communicate with their virtual neighbors by sending requests in any of the γ embedding.

New routing table entries are added by routing for a suitable virtual overlay key, as done in [5] for the tunnel discovery. However, after the routing terminates, the discovered nodes send back their return addresses rather than taking the routing path as a new tunnel. In this manner, the length of routes between virtual overlay neighbors only depends on the trees and does not increase over time. The exact nature of the neighbor discovery, the routing algorithm $\mathbf{R}_{content}$, and the stabilization of the virtual overlay depend on the specifications of the DHT.

Kademlia: In our evaluation in Sections 6 and 7, we utilize a highly resilient recursive Kademlia [23]. In Kademlia, a node selects a *Kademlia identifier* $ID(v)$ uniformly at random in the form of a 160-bit number. The distance between identifiers is equal to their XOR. Nodes maintain many redundant (virtual) overlay connections to increase the resilience. More precisely, each node v keeps a *routing table* of k -buckets. The j -th bucket contains up to k addresses of nodes u so that the common prefix length of $ID(v)$ and $ID(u)$ is j . Maintaining more than 1 neighbor per common prefix length increases the robustness to failures and possibly even to attacks due to the existence of alternative connections.

Based on such routing tables, efficient and robust content discovery is possible. Files are indexed by keys corresponding to the hash of their content, i.e., the algorithm $\mathbf{AdGen}_{content}$ for the generation of file addresses is a hash function. A node u requesting a file with key f looks up the closest nodes v_1, \dots, v_α to f in its routing table in terms of virtual overlay coordinates. Then, u routes for each v_i in the τ trees. Upon receiving one of the messages, v_i returns f via the same route if in possession of f . Otherwise, v forwards the message to the overlay neighbor closest to f , again using tree routing, and returns an acknowledgement message to u . If a node on the route has already received

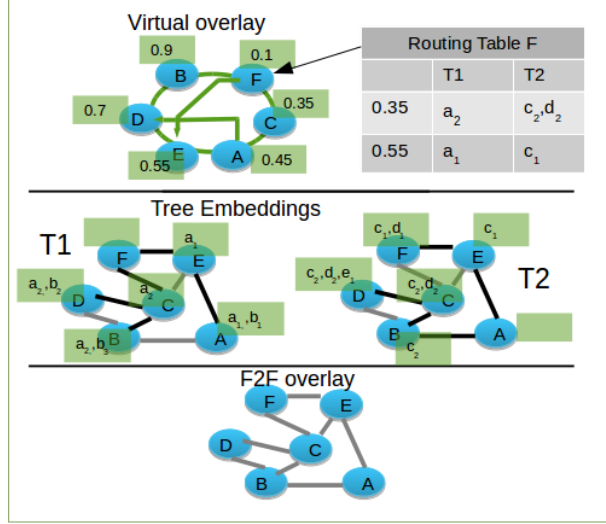


Figure 3: Layers of VOUTE: 1) F2F overlay as restricted topology, 2) Tree embeddings $T1$ and $T2$ offer addressing for messaging, 3) Virtual overlays with tree addresses offer content sharing DHT routing based on tree addresses

the message via a parallel query, it returns a backtrack message such that the predecessor can contact a different node. Similarly, if a node does not receive an acknowledgement from its overlay neighbor in time, it selects an alternative node from its routing table if virtual neighbors closer to f than u exist.

Similarly, stabilization is realized in the same reactive manner as in the original Kademlia. Whenever a node successfully sends a message to an overlay neighbor, this neighbor returns an acknowledgement containing updated return addresses if any coordinates were changed. If a node in the routing table cannot be contacted, the node removes the neighbor from the routing table. Depending on the implementation, it initializes a new neighbor discovery request for the prefix. In addition, suitable neighbors encountered during routing are added to the routing table.

We have now presented the essential components of our design. In the following, we evaluate our design with regard to our requirements. The different layers of our system are displayed in Figure 3

6 Efficiency and Scalability

In this section, we analyze the efficiency of our scheme with regard to routing complexity, stabilization complexity, and their evolution over time.

6.1 Theoretical Analysis

In the first part of this section, we obtain upper bounds on the expected routing length of the routing algorithms \mathbf{R}^{TD} and \mathbf{R}^{CPL} . The desired upper bound on the routing complexity follows by multiplying this bound for routing in one tree with τ , the number of trees used for parallel routing. Afterwards, we consider the stabilization complexity $CS^{\mathbf{S}}$ of the stabilization algorithm \mathbf{S} consisting of i) the local reconstruction of the trees according to Algorithm 1 and ii) the assignment of new coordinates for the nodes affected by a change topology using the modified PIE embedding.

Routing: We consider both messaging between nodes as well as content discovery in the DHT.

Theorem 6.1. *Let id be a modified PIE embedding on a spanning tree of G generated by Algorithm 1 with parameters γ and q . Furthermore, assume that the diameter of G is $\text{diam}(G) = \mathcal{O}(\log n)$. The expected routing length of Algorithm 2 is at most*

$$\mathbb{E}(R^{TD}) = \mathcal{O}\left(\frac{\gamma}{q} \log n\right) \quad (6)$$

for the routing algorithm \mathbf{R}^{TD} , and

$$\mathbb{E}(R^{CPL}) = \mathcal{O}\left(\left(\frac{\gamma}{q}\right)^2 \log n\right) \quad (7)$$

for \mathbf{R}^{CPL} .

For the proof, we first show Lemma 6.2, which bounds the expected level of a node in trees constructed by Algorithm 1. More precisely, we prove that the expected level of a node in any tree constructed by Algorithm 1 is increased by at most a constant factor in comparison to a breath-first-search.

Lemma 6.2. *Let T be any of the γ trees constructed by Algorithm 1 and r the root of T . Furthermore, denote by $sp_r(v)$ the length of the shortest path from v to r , and let $L_T(v)$ be the level of v in T . Then the expected value of $L_T(v)$ is bound by*

$$\mathbb{E}(L_T(v)) \leq sp_r(v) \cdot \left(1 + \frac{\gamma}{q}\right). \quad (8)$$

Proof. We first give an upper bound on the expected number of rounds until a node v accepts an invitation for T after receiving the first invitation. Afterwards, we show Eq. 8 by induction.

In the first step, we denote the number of rounds until acceptance by Y . In order to derive an upper bound on $\mathbb{E}(Y)$, we assume that v does not receive any invitation that it can immediately accept, i.e., an invitation from neighbors u

with minimal parent count $pc(u)$. Thus, v accepts one invitation with probability q in each round. In the worst case, the γ -th accepted invitation is for tree T . The number of rounds thus corresponds to the sum of γ identically distributed geometrically distributed random variables X_1, \dots, X_γ . Here, X_i is the number of trials until the first success of a sequence of Bernoulli experiments with success probability q , i.e., the number of rounds until an invitation is accepted. The random variable $X = X_1 + \dots + X_\gamma$ describes the number of trials until the γ -th success and presents an upper bound on the expected number of rounds until acceptance of an invitation for tree T . We hence derive an upper bound on $\mathbb{E}(Y)$ by

$$\mathbb{E}(Y) \leq \mathbb{E}(X) = \sum_{i=1}^{\gamma} \mathbb{E}(X_i) = \gamma \mathbb{E}(X_1) = \frac{\gamma}{q}. \quad (9)$$

In the second step, we apply induction on $l = sp_r(v)$. Note that the level of a node in the tree is at most the number of rounds until an invitation is accepted from the start of the protocol. For $l = 1$, the node v receives an invitation from r at round 1 of the protocol because v is a neighbor of the root node. In expectation, v joins T at round at most $1 + \mathbb{E}(Y) \leq 1 + \frac{\gamma}{q}$, which shows the claim for $l = 1$. Now, we assume Eq. 8 holds for $l - 1$ and show that then it also holds for l . The number of rounds Z until the node v at level l accepts an invitation in tree T is the sum of Z_1 , the number of rounds until the first invitation is received, and Z_2 the number of rounds v accepts after receiving the first invitation. v is the neighbor of a node w with $sp_r(w) = l - 1$ and receives an invitation from w one round after w joined T . So, Z_1 is bound by our induction hypothesis, and Z_2 is equal to Y and hence bound by Eq. 9. As a result,

$$\mathbb{E}(Z) = \mathbb{E}(Z_1) + 1 + \mathbb{E}(Z_2) \leq (l - 1) \cdot \left(\frac{\gamma}{q} + 1 \right) + 1 + \frac{\gamma}{q} + 1 = l \cdot \left(\frac{\gamma}{q} + 1 \right),$$

and hence indeed Eq. 8 holds. \square

Based on Lemma 6.2, we now prove Theorem 6.1. The idea of the proof is to bound the routing length by a multiple of expected level of a node.

Proof. We consider the diversity measure δ_{RP-TD} first and then δ_{RP-CPL} .

For δ_{RP-TD} , the claim follows directly from Lemma 6.2 and Theorem 4.3 in [8]. More precisely, the expected level of a node is at most $\mathcal{O}\left(\frac{\gamma}{q} \log n\right)$ assuming a diameter and hence maximal distance to the root of $\mathcal{O}(\log n)$. Recall that the distance $\delta_{TD}(id(s), id(e))$ of two nodes s and e corresponds to the length of the shortest path between them in the tree and is an upper bound on the routing. Now, by Eq. 1, the sum of the length of the two coordinates is an upper bound on $\delta_{TD}(id(s), id(e))$. As the length of a coordinate is equal to the level of the corresponding node in the tree, we indeed obtain

$$\mathbb{E}(R_{s,e}^{TD}) \leq \mathbb{E}(\delta_{TD}(id(s), id(e))) \leq \mathbb{E}(L_T(s)) + \mathbb{E}(L_T(e)) = \mathcal{O}\left(\frac{\gamma}{q} \log n\right). \quad (10)$$

The last step follows from Lemma 6.2. Eq. 6 follows because Eq. 10 holds for all source-destination pairs (s, e) .

In contrast, the proof for the common prefix length based similarities cannot build on previous results. Note that the change of the distance function does not affect the existence of a path with expected length at most $\mathbb{E}(L_T(s)) + \mathbb{E}(L_T(e))$ between source s and destination e in the tree. However, the routing might divert from that path when discovering a node with a longer common prefix length but at a higher depth. For this reason, the sum of the expected levels is not an upper bound on the routing length. Rather, whenever a node with a longer common prefix length is contacted, the upper bound of the remaining number of hops is reset to the expected level of that node in addition to the level of e . In the following, we show that such a reset increases the distance in tree by less than $\frac{\gamma}{q}$ on average. The claim then follows because the number of resets is bound by the expected level of the destination. Eq. 7 follows by multiplication of the increased distance per reset and the number of resets.

More precisely, let X_i give the tree distance between the i -th contacted node v_i and the target e . Again, we cannot use the traditional approach for deriving the routing length because X_i is not monotonously decreasing. Rather, we need to bound the number of times Z_1 that X_i increases and the expected amount of increase Z_2 . Thus, the routing length $R_{s,e}^{CPL}$ from a source node s to e is bound by

$$\mathbb{E}(R_{s,e}^{CPL}) \leq \mathbb{E}(L_T(s)) + \mathbb{E}(L_T(e)) + \mathbb{E}(Z_1)\mathbb{E}(Z_2). \quad (11)$$

The number of times Z_1 the common prefix length can increase is bound by the length of the target's coordinate and hence its level in T . So by Lemma 6.2,

$$\mathbb{E}(Z_1) \leq \mathbb{E}(L_T(e)). \quad (12)$$

The tree distance X_i is potentially increased whenever a node with a longer common prefix length is contacted. Yet, an upper bound on the expected increase is given by the difference in the levels $L_T(v_i)$ and $L_T(v_{i+1})$ minus 1 due to the increased common prefix length. Note that v_i and v_{i+1} are neighbors and hence the length of their shortest path to the root differs by at most 1. Lemma 6.2 thus provides the desired bound on $\mathbb{E}(Z_2)$

$$\mathbb{E}(Z_2) \leq \mathbb{E}(L_T(v_i) - L_T(v_{i+1})) - 1 = \frac{\gamma}{q}. \quad (13)$$

The desired bound can now be derived from Lemma 6.2, Eqs. 11, 12, and 13 under the assumption that the diameter of the graph and hence all shortest paths to the root scale logarithmically, i.e.,

$$\mathbb{E}(R_{s,e}^{CPL}) \leq \mathbb{E}(L_T(s)) + \mathbb{E}(L_T(e)) + \mathbb{E}(L_T(e)) \frac{\gamma}{q} = \mathcal{O} \left(\left(\frac{\gamma}{q} \right)^2 \log n \right). \quad (14)$$

As for the first part, Eq. 7 follows because Eq. 14 holds for all pairs (s, e) . \square

The bounds for a virtual overlay lookup based on routing algorithm $\mathbf{R}_{content}$ follow directly from the fact that a DHT lookup requires $\mathcal{O}(\log n)$ overlay hops with each hop corresponding to one route in the network embedding.

Corollary 6.3. *If the DHT used for the virtual overlay offers logarithmic routing, the communication complexity of routing algorithm $\mathbf{R}_{content}$ is*

$$\mathbb{E}(DHT^{TD}) = \mathcal{O}\left(\frac{\gamma}{q} \log^2 n\right)$$

for the diversity measure δ_{RP-TD} and

$$\mathbb{E}(DHT^{CPL}) = \mathcal{O}\left(\left(\frac{\gamma}{q}\right)^2 \log^2 n\right)$$

for diversity measure δ_{RP-CPL} .

Stabilization: The stabilization complexity is required to stay polylog in the network size to allow for scalable communication and content addressing. In the following, we hence give bounds for the self-stabilization of the network embeddings, the costs for the virtual overlay follow by considering the maintenance costs for DHT as suggested for general overlay networks and multiplying with the length of the routes between overlay neighbors.

Theorem 6.4. *We assume the social graph G to be of a logarithmic diameter and a constant average degree. Furthermore, we assume the use of a the root election protocol with complexity $\mathcal{O}(n \log n)$. Then stabilization complexity CS^S of the spanning trees constructed by Algorithm 6.2 with parameters γ and q for one topology change is*

$$\mathbb{E}(CS^S) = \mathcal{O}\left(\gamma \frac{\gamma}{q} \log n\right). \quad (15)$$

Proof. We first consider the complexity for one tree. The general result then follows by multiplying with the number of trees γ . When a node joins an overlay with a constant average degree, the communication complexity of receiving and replying to all invitations is constant. For a node departure, we consider non-root nodes and root nodes separately. If a any node but the root departs, the expected stabilization complexity corresponds to the number of nodes that have to rejoin T . This number of nodes is equal to the number of descendants in a tree. Hence, the expected complexity of a departure corresponds to the expected number of descendants. Consider that a node on level l is a descendant of l nodes, so that the expected number of descendants D is given by

$$\mathbb{E}(D) = \frac{1}{|V|} \sum_{v \in V} \mathbb{E}(L_T(v)) = \mathcal{O}\left(\frac{\gamma}{q} \log n\right).$$

If the root node leaves, the spanning tree and the embedding have to be re-established at a complexity of $\mathcal{O}(n \log n)$. As the probability for the root to depart is $1/n$, we indeed have

$$\mathbb{E}(CS^S) = \mathcal{O}\left(\frac{\gamma}{q} \log n\right) + \mathcal{O}\left(\frac{1}{n} n \log n\right) = \mathcal{O}\left(\frac{\gamma}{q} \log n\right).$$

□

We have shown that the complexity of routing, content discovery, and stabilization is bound (poly-)log as required.

6.2 Simulations

In this section, we validate the above bounds and relate them to the concrete communication overhead for selected scenarios. We start by detailing our simulation model and set-up, followed by our expectation, the results and their interpretation.

Model and Evaluation Metrics: In order to evaluate the efficiency, we consider the routing length and the stabilization complexity. We express the stabilization costs in terms of the average number of coordinates that have to be reassigned when a randomly chosen node leaves, i.e., the average number of descendants of a node. The number of messages required for the assigning the new coordinates is at most two per assignment, namely the disconnected node registering at a new parent and receiving a new coordinate. We conducted the study to determine how the number of trees, the tree construction algorithm, and the distance function affect routing and stabilization costs.

We compared our results to those for Freenet, a virtual overlay *VO*, and the original PIE embedding. The virtual overlay *VO* combines the advantages of X-Vine and MCON by using shortest paths as tunnels in a Kademlia overlay like MCON but integrating backtracking in the presence of local optima and shortcuts from one tunnel to another like X-Vine.

Set-up: Due to space constraints, we restrict the presented results to one example network, namely the giant component of a community network from Facebook with 63392 users ⁵.

The spanning tree construction in Algorithm 1 is parametrized by the number of trees $\gamma \in \{1, 2, 3, 5, 7, 10, 12, 15\}$, the acceptance probability $q = 0.5$, and the selection criterion W chosen to be either random selection (denoted *DIV-RAND*) or preference of nodes at a low depth (denoted *DIV-DEP*). In addition, we consider a breadth first search for spanning tree construction (denoted *BFS*). Moreover, we consider the impact of the two distances δ_{TD} (denoted *TD*) and δ_{CPL} (denoted *CPL*). The length of the return addresses was set to $L = 128$

⁵<https://konect.uni-koblenz.de/networks/facebook-wosn-links>

and the number of bits per element was $b = 128$, all $\tau = \gamma$ embeddings were considered for routing.

For the virtual overlay used for content addressing, we chose a highly resilient recursive Kademlia [23] with bucket size 8 and $\alpha \in \{1, 3\}$ parallel look-ups. Because routing table entries are not uniquely determined by Kademlia identifiers, the entries were chosen randomly from all suitable candidates.

We parametrized the related approaches as follows. For simulating Freenet, we executed the embedding for 6,000 iteration as suggested in [24] and then routed using a distance-directed depth-first search based only on the information about direct neighbors. The routing and stabilization complexity of the original PIE embedding is equal to the respective quantities of our algorithm for $\gamma = 1$, the distance function δ_{TD} and routing without the use of backtracking. In order to better understand the results of the comparison, we simulate the virtual overlay *VO* using the same Kademlia overlay as for our own approach but replacing the tree routing by tunnels corresponding to the shortest paths between overlay neighbors. So, we parametrized the related approaches by either using the proposed standard parameters or selecting parameters that are suitable for comparison because they corresponds to the same degree of redundancy as the parametrization of our own approach.

All results were averaged over 20 runs. They are displayed with 95% confidence intervals. Each run consisted of 100,000 routing attempts for a randomly selected source-destination pair.

Expectations: We expect that the routing length decreases with the number of embeddings, because the number of available routes and thus the probability to discover the shortest route in one embedding increases. In general, the routing length is directly related to the tree depth and should thus be lower for *BFS* and *DIV-DEP*.

Similarly, we expect a higher stabilization overhead for trees of a higher depth as the expected number of descendants per node increases. Thus, the number of nodes that need to select a new parent should be higher for *DIV-RAND* than for *DIV-DEP* and *BFS*.

In comparison to the existing approaches, our approach should enable shorter routes between pairs of nodes than both Freenet and VO. As shown above, we achieve a routing complexity of $\mathcal{O}(\log n)$ whereas the related work achieves at best routes of polylog length. However, our routes for content discovery should be slightly longer than in VO. VO utilizes the same DHT routing but uses shortest paths rather than the longer tree routes.

Results: The impact of the three parameters, number of trees, tree construction, and distance on the routing length confirms our expectations. First, the results indicate that the tree construction, in particular the number of trees, is the dominating factor for the routing length. So, the routing length decreased considerably if multiple embeddings were used because the shortest route in any of the trees was considered. Second, preferring parents closer to the root, i.e.,

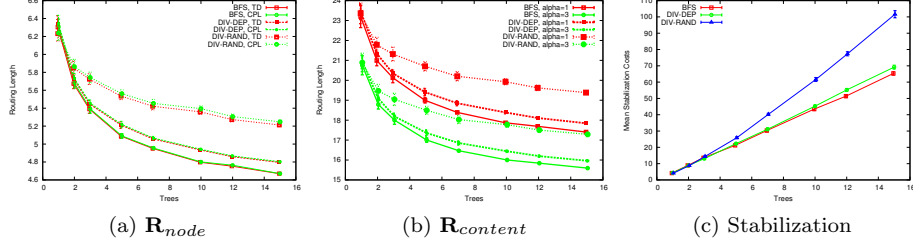


Figure 4: Impact of number of embeddings γ , tree construction, and distance function on routing length for a) tree routing and b) Kademlia lookup with degree of parallelism α ; related approaches result in routing lengths of 14 (virtual overlay *VO*) and close to 10,000 (Freenet), and c) stabilization overhead

using *BFS* or *DIV-DEP*, produced shorter routes in the tree and hence reduced the routing length. Third, in comparison to the tree construction, the choice of a distance function had less impact. For *BFS* or *DIV-DEP*, the advantage of *TD* over *CPL* was barely noticeable, whereas the difference for *DIV-RAND* was still small but noticeable. In order to understand this difference, note that *CPL* is expected to lead to longer routes. The reason for the longer routes lies in forwarding the request to neighbors at a higher depth, which might have a long common prefix but are nevertheless at a higher distance from the destination due to their depth. For *BFS* or *DIV-DEP*, the difference of the depth of neighbors was generally small because neighbors at a lower depth were preferably selected as parents. In contrast, *DIV-DEP* allows for larger differences in depth. Hence there is a higher probability to increase the tree distance by selecting a neighbor with a longer common prefix length but at a high depth. All in all, the routing length varied between 4.67 (*BFS*, $\gamma = 15$, *TD*) and 6.24 (*DIV-RAND*, $\gamma = 1$, *CPL*) hops, as displayed in Figure 4a. In summary, the use of multiple embeddings indeed reduced the routing length considerably.

The performance of the DHT lookup in the virtual overlay directly related to the previous results (cmp. Fig. 4b for the distance under *TD*). The overhead for the discovery of a randomly chosen Kademlia ID, stored at the node with the closest ID in the overlay, varied between 15.56 and 24.25 hops in the F2F overlay, at around 4 hops in the virtual overlay.

By Theorem 6.4, the stabilization complexity was expected to increase at most quadratic with the number of trees. Indeed, Figure 4c supports this fact for *DIV-RAND*. The increase for *BFS* and *DIV-DEP* was even only linear and slightly super-linear, respectively. Note that the quadratic increase is due to the raising average depth of additional trees generated by Algorithm 1. With the goal of achieving diverse spanning trees, nodes select parents at a higher depth. However, the average number of descendants increases with the depth, because a node at depth l is a descendant of l nodes. Due to the stabilization complexity corresponding to the number of the departing node’s descendants, the stabiliza-

tion overhead was higher for *DIV-RAND* and *DIV-DEP* than for *BFS*. More precisely, *BFS* constructs all γ trees independently, so that the average depth of each tree is independent of the number of trees. The stabilization complexity per tree thus remains constant. *DIV-DEP*, aiming to balance diversity and short routes, causes stabilization overhead between the two former approaches, but performed closer to *BFS* (this similarity also held for the routing length). More concretely, the average stabilization overhead for a departing node was slightly below 4.5 for a single tree. For $\gamma = 15$ it increased to 65 (*BFS*), 69 (*DIV-DEP*), and more than 101 (*DIV-RAND*). In contrast to a complete re-computation of the embedding requiring at least $n = 63392$ messages, the stabilization overhead is negligible.

For the related approaches, we found a routing length of 9403.1 for Freenet, 16.11 for VO with $\alpha = 1$, and 14.07 for VO with $\alpha = 3$. Furthermore, the shortest paths are on average of length 4.31, meaning that our routing length of 4.67 is close to optimal. So, routing between nodes in the tree required less than half the overhead of state-of-the-art approaches. Routing in the virtual overlay, requiring at best less than 16 hops in our scheme, was slightly more costly in our approach than in VO due to the inability of the tree routing to guarantee shortest paths between virtual neighbors.

A straight-forward comparison of the stabilization overhead was not possible. Since Freenet stabilizes periodically, there is no overhead directly associated with a leaving node. In case of virtual overlays, VO uses flooding for stabilization, which is clearly more costly. Other overlays such as X-Vine use less costly stabilization but stabilization and routing overhead are unstable and increase over time as shown in [6], so that it is unclear which state of the system should be considered for a comparison. In order to nevertheless give a lower bound on the stabilization overhead, we computed the number of tunnels that needed to be rebuild in VO. On average, 477.35 tunnels corresponding to shortest paths were affected by a departing node. If a tunnel is repaired by routing in the Kademlia overlay like in X-Vine, the stabilization overhead per tunnel corresponds to routing a request and the corresponding reply, i.e., for tunnels corresponding to shortest paths at least $2 \cdot 14 = 28$ messages, resulting in a lower bound on more than 10,000 messages per node departure. The above stabilization algorithm is unable to maintain short routes, such that the actual overhead of stabilization in virtual overlay is even higher than the above lower bound.

Discussion: Our simulation study validates the asymptotic bounds. Indeed, the routing length and thus the routing complexity for messaging is very low, improving on the state-of-the-art by more than a factor of 3. The stabilization complexity is similarly low if the number of trees is not too high. Even for $\gamma = 15$ trees, the number of involved nodes is generally well below 100, which still improves upon virtual overlays such as VO, the most promising state-of-the-art candidate. Only content discovery in form of a DHT lookup was slightly more costly in our approach than in VO, which we consider acceptable given the considerable advantage with regard to all other metrics.

We have considerably improved the efficiency of F2F overlays. In the following, we show that we also mitigated their vulnerability to failures and attacks.

7 Robustness and Censorship-Resistance

In this section, we consider the robustness and resilience to censorship of VOUTE. Note that the evaluation of the censorship-resilience requires a specification of the modified stabilization algorithm \mathbf{S}' , which refer to as *attack strategy* in the following. After deriving two attack strategies, we subsequently present our theoretical and simulation-based evaluation.

We express our results in terms of node coordinates and distances δ_{TD} and δ_{CPL} rather than the corresponding diversity measures. The use of distances simplifies the notation as we do not need to apply a hash cascade for the comparison of coordinates and return addresses. As the routing paths are chosen identical for both coordinates and return addresses, the results are equally valid for return addresses.

7.1 Attack Strategy

We first describe our attack strategies and then comment on additional strategies and our reasons on selecting In order to model secure and insecure root selection protocols, we consider two realizations of *ATT-RAND* and *ATT-ROOT*. In the following, assume that one attacker node has established x links to honest nodes and now aims to censor communication.

For secure spanning trees, the adversary A is unable to manipulate the root election. Nevertheless, A can manipulate the subsequent embedding. The attack strategy *ATT-RAND* assigns each of its children a different random prefix rather than the correct prefix. In this manner, routing fails because nodes in the higher levels of the tree do not recognize the prefix. So, the impact of the attack is increased in comparison to a random failure.

In contrast, if the adversary A can manipulate the root election protocol, *ATT-ROOT* manipulates the root election in all spanning trees such that A becomes the root in all trees. Under the assumption that the root observes the maximal number of requests, the attack should result in a high ratio of failed requests.

Now, we shortly comment on some further attack strategies we choose not to implement and give reasons for our decision not to do so.

First, note that in the original PIE embedding, assigning the same coordinate to two children is another attack strategy. In contrast to the above strategy, the routing can then fail even if the attacker is not involved in forwarding the actual request because the node coordinates are not unique and thus the request might end up at a different node than the receiver. In the modified embedding, the child decides on the last element of the coordinate. Hence, the attacker can only assign a node w the coordinate of another node v as a prefix, so that

the two nodes appear to be close but are indeed not. However, upon realizing that w does not offer a route to v , the routing algorithm backtracks, so that this attack strategy merely increases the routing complexity but not the success ratio. Thus, we do not consider it here.

Second, recall from Section 3 that the attacker can also generate an arbitrary number of identities whereas the above attack strategies only rely on one identity. In the following, we argue that without additional knowledge, the use of additional identities in the tree does not improve the strength of the attack.

ATT-RAND actually simulates different virtual identities by providing fake distinct prefixes to all children. Indeed, in practice, it might be wise to indeed use distinct physical nodes because it minimizes the risk of detection if two neighbors realize that they are connected to the same physical node but received different prefixes.

For *ATT-ROOT*, the attacker might have to create (virtual) identities in order to manipulate the root election. As soon as A is the root in each tree, multiple identities could be used to provide prefixes of different lengths. However, if a neighbor u of A receives a long prefix from A , there is a high chance that u and potential descendants of u choose different parents seemingly closer to the root. Thus, in expectation a large number of nodes joins those subtrees rooted at a neighbor of A with a short prefix. As routing within such a subtree does not require to forward a request from A , A 's impact is likely to be reduced. Hence, without concrete topology knowledge, the insertion of additional virtual identities (corresponding to prefixes of different lengths) does usually not present an obvious advantage for A .

7.2 Theoretical Analysis

We present two theoretical results in this section. First, we characterize the backtracking algorithm more closely. Second, we show that the censorship-resistance is improved by using the distance δ_{CPL} rather than δ_{TD} .

Throughout this section, let \mathbf{R}^{TD} and \mathbf{R}^{CPL} denote Algorithm 2 with distance δ_{TD} and δ_{CPL} , respectively. Furthermore, let \mathbf{GR}^{TD} and \mathbf{GR}^{CPL} denote the corresponding standard greedy routing algorithms, which terminate in local optima with regard to the distance to the destination's coordinate. Let $Succ^{\mathbf{R}}$ denote the success ratio of a routing algorithm \mathbf{R} . We are considering the success ratio for one embedding. The overall success ratio is improved as it is the combined success ratio of all embeddings.

Lemma 7.1. *We have that*

$$\begin{aligned}\mathbb{E}\left(Succ^{\mathbf{R}^{TD}}\right) &\geq \mathbb{E}\left(Succ^{\mathbf{GR}^{TD}}\right) \\ \mathbb{E}\left(Succ^{\mathbf{R}^{CPL}}\right) &\geq \mathbb{E}\left(Succ^{\mathbf{GR}^{CDF}}\right).\end{aligned}\tag{16}$$

Furthermore, Algorithm 2 is successful if and only if there exists a greedy path of responsive nodes according to its distance metric δ_X .

Proof. Eq. 16 follows because Algorithm 2 is identical to the standard greedy algorithm until the latter terminates. Then, Algorithm 2 continues to search for an alternative, possibly increasing the success ratio.

For the second part, recall that a greedy path is a path $p = (v_0, \dots, v_l)$ such that the distance to the destination v_l decreases in each step, i.e., $\delta(id(e), id(v_i)) < \delta(id(e), id(v_{i-1}))$ for all $i = 1..l$ and a distance δ . Assume Algorithm 2 does not discover a route from the source $v_0 = s$ and $v_l = e$ despite the existence of a greedy path $p = (v_0, v_1, \dots, v_{l-1}, v_l)$ of responsive nodes. Let V_R be the set of nodes that forwarded the request according to Algorithm 2, and let $j = \max\{i : v_i \in V_R\}$. Then the neighbor of v_{j+1} did not receive the request despite being closer to e than v_j . Though v_j might have a neighbor w closer to e than v_{j+1} , the request is backtracked to v_j if forwarding to w does not result in a route to the destination. Routing only terminates if either a route is found or v_j has forwarded the request to all closer neighbors, including v_{j+1} . Thus, Algorithm 2 cannot fail if a greedy path exists. In contrast, if there are not any greedy paths from s to e , any path $p = (v_0, v_1, \dots, v_{l-1}, v_l)$ with $v_0 = s$ and $v_l = e$ contains a pair (v_{i-1}, v_i) with $\delta(id(e), id(v_i)) \geq \delta(id(e), id(v_{i-1}))$. Thus, Algorithm 2 does not forward the request to v_i and hence does not discover a path from s to e . It follows that indeed Algorithm 2 is successful if and only if a greedy path of responsive nodes exists. \square

Now, we use Lemma 7.1 to show that using a common prefix length based distance generally enhances the censorship-resistance.

Theorem 7.2. *Let A be an attacker applying either ATT-RAND or ATT-ROOT. Then for all distinct nodes $s, e \in V$*

$$Succ^{\mathbf{R}^{CPL}}_{s,e} = 0 \implies Succ^{\mathbf{R}^{TD}}_{s,e} = 0, \quad (17)$$

i.e., if \mathbf{R}^{CPL} does not discover a route between s and e , then \mathbf{R}^{TD} does not discover a route. However, the converse does not hold. In particular,

$$\mathbb{E} \left(Succ^{\mathbf{R}^{CPL}} \right) \geq \mathbb{E} \left(Succ^{\mathbf{R}^{TD}} \right). \quad (18)$$

Proof. We prove the claim by contradiction. Assume that there is pair s, e such that the algorithm \mathbf{R}^{TD} terminates successfully while \mathbf{R}^{CPL} does not. Let $p = (v_0, v_1, \dots, v_l)$ with $v_0 = s$ and $v_l = e$ denote the discovered route. By Lemma 7.1, p is a greedy path for distance δ_{TD} but not for δ_{CPL} . In other words, there exists $0 \leq i < l$ such that i) $\delta_{TD}(id(v_{i+1}), id(e)) < \delta_{TD}(id(v_i), id(e))$ and ii) $\delta_{CPL}(id(v_{i+1}), id(e)) \geq \delta_{CPL}(id(v_i), id(e))$. By the definitions of both distances in Eq. 1 and Eq. 2, this implies that $cpl(id(v_{i+1}), id(e)) < cpl(id(v_i), id(e))$ and $|id(v_{i+1})| < |id(v_i)|$. In other words, v_{i+1} 's coordinate has a lower common prefix length to $id(e)$ and is shorter than $id(v_i)$. The right side of Figure 5 displays an example.

We base our contradiction upon the following observation concerning routes in trees. Consider the *tree route* between two nodes, i.e., the path between

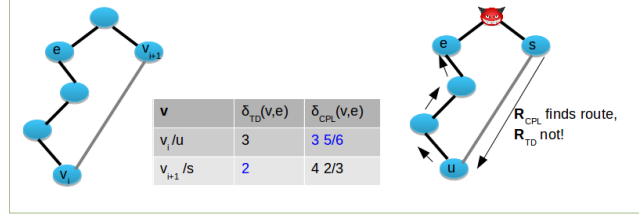


Figure 5: Illustrating the proof of Theorem 7.2: left: v_{i+1} is closer to destination e than v_i for distance δ_{TD} but not for δ_{CPL} ; right: pair (s, e) for which \mathbf{R}^{CPL} is successful as s forwards to u , but \mathbf{R}^{TD} is not successful because s forwards to the attacker.

them using only tree edges. Along the tree route, the common prefix length stays constant until the least common ancestor is reached and then increases. Now, if $id(v_{i+1})$ has a shorter common prefix with $id(e)$ than $id(v_i)$, v_{i+1} is not contained in the tree route. Furthermore, as the routing algorithm \mathbf{R}^{CPL} does not successfully discover a route, the attacker has to control one node on the tree route.

We can use the above observation to establish contradictions for both *ATT-RAND* and *ATT-ROOT*. Note that if the common prefix length decreases when forwarding to v_{i+1} , we need to have $cpl(id(v_i), id(e)) > 0$. For the attack strategy *ATT-RAND*, the attacker on the tree route is either an ancestor of v_i or of e . However, the attacker replaces the prefixes of all its children and hence descendants, so that the perceived common prefix length of v_i ' and e 's coordinates should be 0 unless there exists an attacker-free tree route. This is a clear contradiction. Similarly, if $cpl(id(v_i), id(e)) > 0$, v_i and e have a common ancestor aside from the root. In particular, the tree route does not pass the root. When applying *ATT-ROOT*, the only attacker is the root, which again contradicts that there is an attacker on the tree route. Thus, we have shown by contradiction that \mathbf{R}^{TD} only succeeds if \mathbf{R}^{CPL} does.

Thus, we have shown that indeed Eq. 17 holds. Eq. 18 is a direct consequence as it averages over all source-destination pairs and systems. It remains to show that the converse of Eq. 17 does not hold. In other words, there exist instances when \mathbf{R}^{CPL} terminates successfully while \mathbf{R}^{TD} fails. We display such an example in Figure 5. \square

While we can show that our enhancements are indeed enhancements, our theoretical analysis does not provide any absolute bounds on the success ratio. In particular, we cannot compare our success ratio to that of virtual overlays.

7.3 Simulations

We utilized the simulation model and set-up from Section 6 for evaluating the efficiency and extended it to include the methodology for robustness and

censorship-resistance. So, we simulate the robustness of an overlay by subsequently selecting random failed nodes. In each step, we select a certain fraction of additional failed nodes and then determine the success ratio. Furthermore, we evaluate attacks using the two attack strategies *ATT-RAND* and *ATT-ROOT* described above. We first establish the overlay applying the respective attack strategy and then execute the routing for randomly selected source-destination pairs of responsive nodes.

We compare our results to the virtual overlay VO, described in Section 6. Our attacker on VO does not manipulate the tunnel establishment but merely drops requests. Recall that routing in VO relies on a Kademlia DHT such that neighbors in the DHT communicate via a tunnel of trusted links. The routing between two DHT neighbors thus fails if the attacker is contained in the tunnel. However, if routing between two overlay neighbors fails, the startpoint of the failed tunnel can attempt to select a different overlay neighbor as long as it has one neighbor closer to the destination. We further enhance the success ratio of VO by optionally allowing backtracking in the DHT. In addition, we also allow for shortcuts, i.e., rather than following the tunnel to its endpoint, nodes on the path can change to a different tunnel with an endpoint closer to the destination. Thus, we maximize the chance of successful delivery in VO by backtracking and shortcuts in addition to the use of non-strategic attacker.

Set-up: We used the embedding and routing algorithms as parametrized in Section 6.

In order to evaluate the robustness, we removed up to 50% of the nodes in steps of 1%. During the process of removing nodes, individual nodes inevitably became disconnected from the giant component, so that routing between some pairs was no longer possible. For this reason, we only considered the results for source-destination pairs in the same component. Our results are presented for 1, 5, and 15 trees only.

The number of edges x controlled by the adversary A were chosen as $x = 2^i \times \lceil \log_2 n \rceil$ with $0 \leq i \leq 6$ and $\lceil \log_2 n \rceil = 16$. So, up to 1,024 attacker edges were considered. In particular, $x = 1024 > \frac{\sqrt{n}}{\log n}$, a common asymptotic bound on the number of edges to honest nodes considered for Sybil detection schemes [25]. For quantifying the achieved improvement, we compared our approach to the resilience of the original PIE embedding and routing, i.e., 1 tree, δ_{TD} , and no backtracking.

For VO, we used a degree of parallelism of $\alpha = 1$. Since backtracking was applied, all values of $\alpha > 0$ resulted in the same success ratio, because regardless of the value of α , the routing succeeded if and only if a greedy path in the virtual overlay existed. Thus, restricting our evaluation to $\alpha = 1$ did not impact our results with regard to the success ratio.

We averaged the results over 20 runs with 10,000 source-destination pairs each. Results are presented with 95% confidence intervals.

Expectations: We expect that the use of backtracking already increases the success ratio considerably for $\gamma = 1$. However, for large failure ratios or a large number of attacker edges, the single-connected nature of the tree should result in a low success ratio. By using multiple trees, we expect to further increase the success ratio until close to 100% of the paths correspond to a greedy path and hence a route in at least one embedding.

For the robustness to failures, the original distance function *TD* should result in a higher success ratio than *CPL* because of its shorter routes, as seen in Section 6.2, and thus lower probability to encounter a random failed node. However, by Theorem 7.2, *CPL* increases the success ratio in contrast to the original distance.

Our first attack strategy, *ATT-RAND*, should not have a strong impact as the fraction of controlled edges is low and the attacker usually does not have an important position in the trees. In contrast, we expect many requests to be routed via the root, so that at least for a low number of trees, *ATT-ROOT* should be an effective attack strategy.

In comparison, our attack on VO does not enable the attacker to obtain a position of strategic importance, so that the impact of the attack should be much less drastic than *ATT-ROOT*. However, communication between DHT neighbors relies on one tunnel whereas tree embeddings provide multiple routes. Thus, when using multiple diverse trees, we expect our approach to be similarly effective as VO, possibly even more effective.

Results: While the results verified our expectations with regard to the advantage of the distance *TD* for random failures and of *CPL* for attacks, the observed differences between the two distances were negligible, i.e., less than 0.1%. Hence, we present the results for *CPL* in the following with the exception of the results for the original PIE embedding.

We start by evaluating the robustness to random failures. The results, displayed in Figure 6a, indicate that the use of multiple embeddings considerably increased the robustness. The success ratio for $\gamma = 1$ was low, decreasing in a linear fashion to less than 30% for a failure ratio of 50%. In contrast, for $\gamma = 15$, the success ratio exceeded 90%. Though the number of embeddings was the dominating factor, the tree constructing algorithm also strongly influenced the success ratio. For $\gamma > 1$, aiming to choose distinct parents improved the robustness to failures because of the higher number of distinct routes. For example, when routing in 5 parallel embeddings, the success ratio was above 80% for *DIV-RAND*. In contrast, *BFS* had a success ratio below 70%. In summary, the robustness to failures was extremely high for multiple embeddings, enabling a success ratio of more than 95% for up to 20% failed nodes. The robustness was further increased by using *DIV-RAND* or *DIV-DEP* rather than *BFS*, showing that even such relatively simple schemes can achieve a noticeable improvement.

Now, we consider the censorship-resistance for $x = 16$ attacking edges, as displayed in Figure 6b. If the adversary *A* was unable to manipulate the root selection, the success ratio was only slightly below 100%. Even if $\gamma = 1$, more than

99.5% of the routes were successfully discovered. The high resilience against *ATT-RAND* was to be expected, considering that the attack was only slightly more severe than failure of one random node. If the attacker was able to become the root in all trees, the success ratio dropped to about 93% for $\gamma = 1$. However, with multiple trees, the ratio of *ATT-ROOT* was close to 100%. The impact of the tree construction was small but noticeable. So, *BFS* generally resulted in a slightly lower success ratio. Hence, by using multiple embeddings and backtracking, the resilience to an adversary that can establish only $\lceil \log_2 |V| \rceil$ is such that nearly all routes are successfully found.

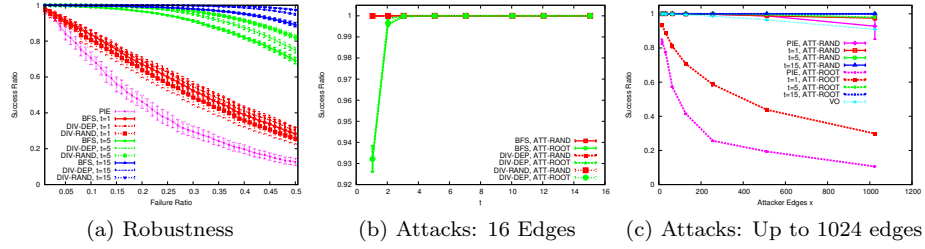


Figure 6: a) Robustness to failures for distance *CPL* and b),c) Censorship-Resistance of tree routing for distance *CPL* to adversaries which are either able to undermine the root election (*ATT-ROOT*) or are unable to do so (*ATT-RAND*) for b) $x = 16$ attacking edges, and c) up to 1,024 attacking edges and tree construction *DIV-DEP*

For an increased number of attacking edges x , the success ratio remained close to 100% when more than one tree was used for routing, as displayed in Figure 6c for *DIV-DEP*. However, for one tree, the success ratio decreased drastically if an attacker could undermine the root selection. For $x = 1024$, i.e., if the attacker controlled edges to roughly 1.7% of the nodes, the success ratio for $\gamma = 1$ decreased to slightly less than 30%. In contrast, if $\gamma = 5$ or $\gamma = 15$, the success ratio was still 97.9 or 99.9%, respectively.

In order to quantify the improvements provided by our resilience enhancements, we compared the results for our approach with the PIE embedding. As can be seen from Figure 6a, the success ratio dropped much more quickly for PIE than for the improved approaches. For an adversary with $x = 16$ connections to honest nodes, PIE suffered from more than twice the numbers of failed requests than the remaining systems (Figure 6c) because it relies on only one tree and does not apply backtracking. When increasing the number of attacker edges, the success ratio dropped further to less than 15% for $x = 1024$. Our approach achieved more than twice the success ratio even for $\gamma = 1$.

In contrast to PIE, VO exhibited a rather high success ratio as displayed in Figure 6c. VO's advantage in contrast to $\gamma = 1$ holds despite VO's longer routes (see Section 6.2). The reason for VO's lower vulnerability lies in the absence of strategic manipulation. While greedy embeddings allow the attacker to assume

an important role, our attacker in VO does not attract a disproportional fraction of traffic. However, establishing multiple trees ensures that the role of the root is effectively mitigated, so that the censorship-resilience of VO is slightly lower than VOUTe’s resilience for 5 or more parallel embeddings.

Discussion: We have shown that multiple embedding and backtracking enable high resilience, outperforming state-of-the-art approaches. Here, we focused on node-to-node communication. While content retrieval results in longer routes, we expect the success ratio to be similar as backtracking in the DHT allows the use of multiple paths. In addition, the number of replicas per content can be adjusted to increase the success ratio.

While Theorem 7.2 shows the advantage of *CPL* in the presence of failures, the actual advantage is negligible, so that it seems more sensible to use the original distance *TD* due to its higher efficiency.

In summary, our enhancements to the robustness and censorship-resistance were both needed and highly effective.

8 Anonymity and Membership-Concealment

We show that our return addresses provide plausible deniability.

Theorem 8.1. *Let u be a local attacker, which is aware only of its direct neighbors N_u in the social graph, and let $y = (y_1, \dots, y_\gamma)$ with routing information $\tilde{k} = (\tilde{k}_1, \dots, \tilde{k}_\gamma)$ be a vector of return addresses generated by Algorithm 3 for the node u_y . Then u cannot identify e_y with absolute certainty using a polynomial-time algorithm A , i.e., $P(A(y, \tilde{k}) = e_y) < 1$ for all return address vectors y . So, we guarantee possible innocence with regard to both sender and receiver anonymity in the absence of identifying side channel information such as timing analysis.*

Proof. We start by considering receiver anonymity. We consider three cases and show for each case that either i) the attacker can determine that the receiver is not a neighbor but cannot infer the coordinate of the actual receiver or ii) the attacker remains uncertain if the receiver is a neighbor or a neighbor’s descendant. We illustrate the cases in Figure 7. Throughout the proof, let $v_i \in N_u$ be the closest neighbor of u to y_i for $i = 1, \dots, \gamma$.

First, assume there exist i, j such that $v_i \neq v_j$. It follows that none of u ’s neighbors is the receiver due to the fact that the receiver can be identified as the closest node to all return addresses. So, u , not being aware of the remaining nodes and their coordinates in the system, cannot identify the receiver.

For the second case, assume that indeed $v_i = v_j$ for all $1 \leq i, j \leq \gamma$ but there exists an i such that $cpl(hc(id(v_i), \tilde{k}_i), y_i) < |id(v_i)|$, i.e., the common prefix length of $id(v_i)$ and the target coordinate is less than the length of v_i ’s coordinate. Then v_i cannot be the receiver because at least the last element in

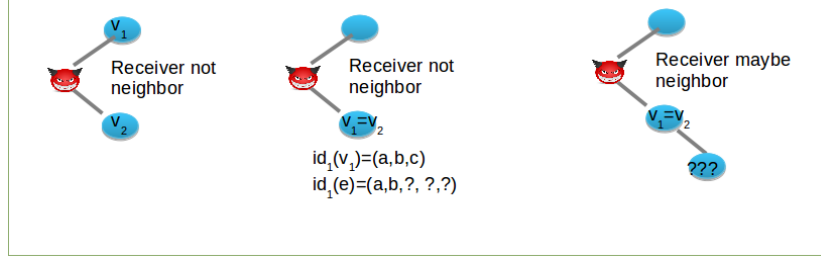


Figure 7: Illustrating the proof of Theorem 8.1: Let v_1 and v_2 be the neighbors with the closest coordinates to the receiver's coordinates $id_1(e)$ and $id_2(e)$ in the first and second embedding, respectively. The attacker can only infer if neighbors are not the receiver e but can not tell if they are. In particular, the attacker A knows the common prefix of the receiver's coordinates and the coordinates of its neighbors but not the remaining elements of the coordinate, as indicated by the ?s in the coordinate. In the first scenario, $v_1 \neq v_2$ shows that the receiver is not a neighbor. In the second scenario, A can infer that the third element of the receiver's coordinate $id_1(e)$ is not c and hence v_1 is not the receiver. In the last scenario, A is unable to tell if a neighbor is indeed the receiver or if a child of the neighbor is the receiver.

the i -th coordinate of v_i does not agree with $id_i(e_y)$. So, again the receiver is not a neighbor of u and hence u is unable to identify e due to its limited view of the overlay.

Third, assume that indeed $v_i = v_j$ for all $1 \leq i, j, \leq \gamma$ and $cpl(hc(id(v_i), \tilde{k}_i), y_i) = |id(v_i)|$ for all i . Then, the node v_i can potentially be the receiver but so can any node w that is a descendant of v_i in all trees. Any return address vector of w would result in the same results as a return address vector of v_i from u 's local point of view. Due to its restricted topology knowledge, u is unaware if such a descendant w exists, and hence can only guess that e_y is the receiver but cannot be certain.

Thus, receiver anonymity follows as the return address does not allow the unique identification of the receiver. Sender anonymity follows analogously as a node can always forward a request from a child. \square

Note that the above proof does not require the application of the hash cascade. However, without the application, an attacker can always infer the common prefix length of two receiver addresses. If we apply the hash cascade, the attacker can only determine the distance of receiver addresses to its own coordinates but might be unable to detect how close two addresses actually are. In this manner, the attacker can only infer very limited topology information. By hiding the topology of the social graph, we prevent the identification of users by comparing a pseudonymous topology to an external social graph.

9 Conclusion

We have introduced a privacy-preserving, efficient, and resilient design for F2F overlays. For this purpose, we have developed an algorithm for the generation of anonymous return addresses. Furthermore, we have designed multiple parallel network embeddings to enable both efficiency and resilience, as validated by an extensive simulation study.

Extending our simulation results, we are currently integrating our algorithms in an existing F2F overlay and have started initial testbed studies to better understand the system and its performance in real environments.

References

- [1] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. Technical report, DTIC Document, 2004.
- [2] Ian Clarke, Oskar Sandberg, Matthew Toseland, and Vilhelm Verendel. Private communication through a network of trusted connections: The dark freenet. *Network*, 2010.
- [3] Eugene Vasserman, Rob Jansen, James Tyra, Nicholas Hopper, and Yongdae Kim. Membership-concealing overlay networks. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 390–399. ACM, 2009.
- [4] Nathan S Evans, Chris GauthierDickey, and Christian Grothoff. Routing in the dark: Pitch black. In *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual*, pages 305–314. IEEE, 2007.
- [5] Prateek Mittal, Matthew Caesar, and Nikita Borisov. X-vine: Secure and pseudonymous routing in dhds using social networks. In *NDSS*, 2012.
- [6] Stefanie Roos and Thorsten Strufe. On the impossibility of efficient self-stabilization in virtual overlays with churn. In *INFOCOM*. IEEE, 2015.
- [7] Robert Kleinberg. Geographic routing using hyperbolic space. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications*. IEEE, pages 1902–1909. IEEE, 2007.
- [8] Julien Herzen, Cedric Westphal, and Patrick Thiran. Scalable routing easy as pie: A practical isometric embedding protocol. In *Network Protocols (ICNP), 2011 19th IEEE International Conference on*, pages 49–58. IEEE, 2011.
- [9] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 173–187. IEEE, 2009.

- [10] Bogdan Popescu. Safe and private data sharing with turtle: friends team-up and beat the system (transcript of discussion). In *Security Protocols*, pages 221–230. Springer, 2006.
- [11] Tomas Isdal, Michael Piatek, Arvind Krishnamurthy, and Thomas Anderson. Privacy-preserving p2p data sharing with oneswarm. In *ACM SIGCOMM Computer Communication Review*, volume 40, pages 111–122. ACM, 2010.
- [12] Nathan S Evans and Christian Grothoff. R5n: Randomized recursive routing for restricted-route networks. In *NSS*, pages 316–321, 2011.
- [13] Andreas Hofer, Stefanie Roos, and Thorsten Strufe. Greedy embedding, routing and content addressing for darknets. In *Networked Systems (Net-Sys), 2013 Conference on*, pages 43–50. IEEE, 2013.
- [14] Jon McLachlan, Andrew Tran, Nicholas Hopper, and Yongdae Kim. Scalable onion routing with torsk. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 590–599. ACM, 2009.
- [15] Atul Singh et al. Eclipse attacks on overlay networks: Threats and defenses. In *IEEE INFOCOM*, 2006.
- [16] Hooman Mohajeri Moghaddam, Baiyu Li, Mohammad Derakhshani, and Ian Goldberg. Skypemorph: Protocol obfuscation for tor bridges. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 97–108. ACM, 2012.
- [17] Christos H Papadimitriou and David Ratajczak. On a conjecture related to geometric routing. In *Algorithmic Aspects of Wireless Sensor Networks*, pages 9–17. 2004.
- [18] Andrej Cvetkovski and Mark Crovella. Hyperbolic embedding and routing for dynamic graphs. In *INFOCOM 2009, IEEE*, pages 1647–1655. IEEE, 2009.
- [19] David Eppstein and Michael T Goodrich. Succinct greedy graph drawing in the hyperbolic plane. In *Graph Drawing*, pages 14–25. Springer, 2009.
- [20] Radia Perlman. An algorithm for distributed computation of a spanningtree in an extended lan. *ACM SIGCOMM Computer Communication Review*, 15(4):44–53, 1985.
- [21] Michael Sirivianos, Dirk Westhoff, Frederik Armknecht, and Joao Girao. Non-manipulable aggregator node election protocols for wireless sensor networks. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, 5th International Symposium on*. IEEE, 2007.
- [22] Michael Brinkmeier, Günter Schäfer, and Thorsten Strufe. Optimally dos resistant p2p topologies for live multimedia streaming. *TPDS*, 20(6), 2009.

- [23] Bernhard Heep. R/kademlia: Recursive and topology-aware overlay routing. In *Telecommunication Networks and Applications Conference (ATNAC), 2010 Australasian*, pages 102–107. IEEE, 2010.
- [24] Oskar Sandberg. Distributed routing in small-world networks. In *ALLENEX*, pages 144–155. SIAM, 2006.
- [25] George Danezis and Prateek Mittal. Sybilinfer: Detecting sybil nodes using social networks. In *NDSS*, 2009.

Appendix

In this section, we show how to obfuscate return addresses in VOUTE further. PPP addresses are then generated by adding an additional layer of symmetric encryption to RAP return addresses generated by Algorithm 3. The idea of the approach is to allow u to determine if the common prefix length of a coordinate is longer than $cpl(cord(y), id(u))$ but not the actual length. For this reason, the additional layer can only be applied when using the common prefix length as a distance. In a nutshell, we generate PPP return address through symmetrically encryption of a RAP return address using key material only known within certain subtrees.

Let $Enc : H \times \mathbb{K}_{Sym} \rightarrow H$ be a semantically secure symmetric encryption function onto h 's image H with keyspace \mathbb{K}_{Sym} . $Dec : H \times Sym \rightarrow H$ denotes the corresponding decryption. For each subtree of the spanning tree, we distribute keys. To achieve that, each internal node w at level l generates a symmetric key $k_l(w)$ by a pseudo-random key generation algorithm $SymGen$. Subsequently, w distributes $k_l(w)$ to all its descendants. In this manner, a node v at level \tilde{l} obtains keys $k_1(v), \dots, k_{\tilde{l}-1}(v)$ such that $k_\lambda(v)$ was generated by v 's ancestor at level λ and forwarded to v along the tree edges. So, $k_\lambda(v)$ is known to all nodes having a common prefix length of at least λ with v . After generating a RAP return address $y = (d_1, \dots, d_L)$, v additionally encrypts the $\lambda + 1$ -th element with the key $k_\lambda(v)$, constructing the return address $y' = (d'_1, \dots, d'_L)$ with

$$d'_j = \begin{cases} Enc(k_{j-1}(v), d_j), & 2 \leq j \leq l \\ d_j, & \text{otherwise} \end{cases} \quad (19)$$

The second case in Eq. 19 treats the first element, which remains unencrypted, and the randomly chosen padding, which does not agree with any coordinate. After generating y' , v publishes y' , the routing information \tilde{k} for generating y and $mac(\mathbb{K}_{MAC}(v), y')$. The pseudo code of the additional encryption is displayed in Algorithm 4.

A third realization \mathbf{R}^{PPP} of the routing algorithm \mathbf{R}_{node} is given by the construction of PPP addresses. During routing, a node u at level l first applies the decryption function to the second to $l + 1$ -th element of the return address

$y' = (d'_1, \dots, d'_L)$. So, v obtains $f(y') = (z_1, \dots, z_{l+1})$ with

$$z_j = \begin{cases} d'_1, & j = 1 \\ \text{Dec}(k_{j-1}(u), d'_j), & \text{otherwise} \end{cases}.$$

Afterwards, u determines $\text{cpl}(f(y'), \text{cash}(\tilde{k}, c))$ for all coordinates c in its neighborhood. Note that $\text{cpl}(f(y'), \text{cash}(\tilde{k}, c))$ is only a lower bound on $\text{cpl}(\text{cord}(y), c) = \text{cpl}(y, \text{cash}(\tilde{k}, c))$ because u is not able to correctly decrypt some elements of y' . Based on the common prefix length, v can evaluate the diversity measure

$$\delta_{PPP-CPL}^u(y', \tilde{k}, c) = \delta_{CPL}(f(y'), \text{cash}(c, \tilde{k})) \quad (20)$$

for the distance δ_{CPL} defined in Eq. 2. In this manner, the node u obtains a set of all neighbors closer to the destination than itself. So, u chooses a random node from this set as the next hop.

We here give a short intuition on why Algorithm 4 indeed generates PPP anonymous return addresses. A formal proof is sketched in Section ???. Let u be a arbitrary node and y' be a return address generated by v , a node at level l_v . If $\text{cpl}(\text{id}(v), \text{id}(u)) = \lambda$, u correctly decrypts the first $\lambda + 1$ elements of y' because $k_i(u) = k_i(v)$ for $i = 1 \dots \lambda$. Due to the semantic security of the symmetric encryption, u cannot infer information about the remaining elements of y from $d'_{\lambda+2}, \dots, d'_{l_u}$ because u does not know $k_i(v)$ for $i > \lambda + 1$. Thus, y' indeed only reveals if a coordinate c shares a longer common prefix with $\text{cord}(y)$ than $\text{id}(v)$.

Algorithm 4 addPPPLayer()

{Input: RAP return address $y = (d_1, \dots, d_L)$ }
{Internal State: Keys $k_1(v), \dots, k_{l-1}(v)$, Enc_{Sym} }
1: **for** $i = 2 \dots l$ **do**
2: $d_j \leftarrow \text{Enc}_{Sym}(k_{j-1}(v), d_j)$ *{Encrypt element j }*
3: **end for**
